Supporting Information

A Feature-aligned Diffusion Model for Controllable Generation of 3D Drug-like Molecules

Hao Lu[1,2], Zhiqiang Wei[1,5], Xiancong Hou[1], Wenzheng Han[1], Hao Liu[1*], Yang Zhang[2,3,4*]

[1]College of Computer Science and Technology, Ocean University of China, Qingdao, 266100, Shandong Province, China.
[2]Department of Computer Science, School of Computing, National University of Singapore, 117417, Singapore.
[3]Cancer Science Institute of Singapore, National University of Singapore, 117599, Singapore.
[4]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 117596, Singapore.
[5]College of Computer Science and Technology, Qingdao University, Qingdao, 266071, Shandong Province, China.

*Corresponding author(s). E-mail(s): liu.hao@ouc.edu.cn; ybfirst@ouc.edu.cn; zhang@nus.edu.sg;

Table of Content

Supporting Texts

Text S1. Geometric and steric evaluation of generated molecules

We performed quantitative assessments of molecular geometry to ensure that the generated molecules maintain chemical plausibility within the binding environment. Following the methodology of Lin et al. [1], we examined bond lengths, bond angles, and steric clashes for the generated compounds. Selected metrics are summarized as follows:

- **JSD_{BL}(↓)** (Jensen–Shannon Divergence of Bond Lengths): This measures the statistical divergence between the bond length distributions of generated molecules and reference molecules. A lower JSD_{BL} indicates that the generated bond lengths are more consistent with those observed in real molecules, reflecting better geometric fidelity.
- **JSD_{BA}(↓)** (Jensen–Shannon Divergence of Bond Angles): This evaluates the divergence in bond angle distributions between generated and reference molecules. A smaller JSD_{BA} suggests that the bond angles in generated molecules better align with natural conformational statistics, implying higher chemical realism.
- **$Ratio_{cca}$(↓)** (Ratio of Cross-Clash Atoms): This metric quantifies atom-level steric clashes between generated molecules and protein atoms. A clash is defined as a van der Waals overlap of ≥ 0.4 Å (Ramachandran et al., [2]). $Ratio_{cca}$ represents the fraction of atoms in a molecule that participate in such clashes — the lower the value, the fewer physically implausible interactions.
- **$Ratio_{cm}$(↓)** (Ratio of Clash Molecules): This measures the percentage of generated molecules that contain at least one steric clash with protein atoms. A lower $Ratio_{cm}$ indicates that more generated molecules are structurally compatible with the binding site.

We compared multiple state-of-the-art baselines against our proposed ExpDiff model using four stringent metrics (as shown in **Table S1**). For geometric accuracy, ExpDiff achieved a bond length divergence of JSD_{BL} = 0.2808 and a bond angle divergence of JSD_{BA} = 0.3886, placing it among the top-performing methods. While not the absolute best across all metrics, ExpDiff clearly outperforms widely used generative approaches. The bond angle distribution likewise ranks among the leading positions relative to current baseline models (**Table S2**). These results indicate that ExpDiff reproduces the fine-grained geometric distributions of real molecules with higher fidelity, thereby mitigating the distortions in bond statistics that often undermine chemical validity. For steric clash analysis, ExpDiff exhibited a very low atom-level clash ratio ($Ratio_{cca}$ = 0.0161), ranking second overall (the lowest value was attained by LIGAN, 0.0096). ExpDiff therefore substantially reduces unphysical overlaps compared with most baselines. This low per-atom clash burden translates into a reduced fraction of clash-containing molecules ($Ratio_{cm}$ = 0.2696), which is likewise the second-lowest after LIGAN (0.0718) and is approximately 60% of the $Ratio_{cm}$ observed for Pocket2Mol (0.4499). Taken together, these findings establish that ExpDiff not only captures the intrinsic geometric regularities of chemical structures but also ensures structural compatibility with the target protein binding site, thereby advancing the field toward practical, structure-based molecular generation.

Text S2. Ablation study

We employed an ablation study to demonstrate the utility of the proposed expert model for feature alignment. That is, the hyperparameter ε in Eq. (13). **Table 3** shows the effect of different alignment factors on the molecular affinity generated by ExpDiff, and the results are based on the average and median of the three metrics Vina score, Vina minimize and Vina dock. The results show that $\text{ExpDiff}_{0.1}$ performs the best in all metrics, achieving the optimal mean of Vina score (-9.096), Vina minimize (-9.579) and Vina dock (-9.920), as well as the corresponding optimal median, suggesting that this alignment factor setting is capable of generating molecules with higher affinity. In contrast, ExpDiff_0 and ExpDiff_1 performed poorly, suggesting that the choice of alignment factor has a significant impact on the quality of molecules generated.

Further, to verify whether there should be different alignment weights at different diffusion steps, we use invariant weights versus variable weights to evaluate the impact during the training of the ExpDiff. **Figure S1a** shows the change of weight values under different weighting strategies during the training process. The constant weights remain constant while the variable weights gradually decrease, indicating the adaptation of the model to the variable weights during the training process. The variation of the position and total loss of atoms from 500,000 to 762,000 training iteration steps is presented in **Figure S1b** and **Figure S1c**. The average loss in the positions of the ligands is 0.705 (variable weights) and 0.703 (constant weights), respectively. The model using variable weights (blue line) is slightly lower than the constant weight model (yellow line) in the range of loss fluctuations, signaling the potential advantage of the variable weight strategy in improving training stability. **Figure. S1c** shows how the total loss varies across different training iteration steps. Similar to the results for the positional loss of atoms, the total loss shows the same trend. **Figure S1d** summarizes the docking results under each configuration. Both the mean scores and the median show better docking results for the model using variable weights.

Additionally, we tested the relationship between adding alignment networks, variable weights, and affinity prediction networks to predict affinity. As shown in **Figure S1e**, the Person correlation coefficient between the affinity prediction network of the model without the alignment module and the Vina dock results is -0.878, while the corresponding values of the model with an alignment factor of 0.1 and the model with variable loss are -0.887 and -0.913, respectively, which shows that the accuracy of the affinity prediction network is improved. Therefore, the method of using variable weights shows remarkable advantages in training the ExpDiff, both in terms of the stability of the loss function and the optimization of the alignment results.

Text S3. Binding free energy via MM/GB(PB)SA

We employed Uni-GBSA [3] to rapidly evaluate the binding free energies between the generated molecules and target proteins. Uni-GBSA provides an automated MM/GB(PB)SA workflow, which enables efficient assessment of protein–ligand binding affinities in the context of virtual screening and molecular optimization. The Crossdocked2020 dataset was used to supply the test protein structures, and all proteins were preprocessed and optimized using PDBFixer [4] prior to energy calculations. For each model, 20 generated molecules were docked into the test proteins and subjected to binding free energy evaluation. It should be noted that several proteins from the dataset produced errors during the calculation and

were therefore excluded. As a result, a total of 91 proteins were successfully analyzed, with their corresponding protein IDs summarized in **Table S5.**

The binding free energy (ΔG) of a protein–ligand complex is evaluated according to Eq. (1) in the MM/GB(PB)SA framework [5,6].

$$\Delta G = \Delta G_{protein-ligand} - \Delta G_{protein} - \Delta G_{ligand} \quad (1)$$

Alternatively, ΔG can be expressed as:

$$\Delta G = \Delta H - T\Delta S \quad (2)$$
$$\Delta H = \Delta E_{MM} + \Delta G_{SOLV} \quad (3)$$

In this context, ΔH represents the enthalpic component, which can be partitioned into the molecular mechanics (MM) energy in the gas phase (ΔE_{MM}) and the solvation contribution (ΔG_{SOLV}), as defined in Eq. (3). The entropic term ($-T\Delta S$) accounts for conformational entropy changes upon ligand association, with T denoting the absolute temperature. Because estimating conformational entropy is computationally demanding, this contribution is frequently omitted. As a result, the calculated effective free energy is generally adequate for assessing relative binding affinities among structurally related ligands [7]. The MM energy term ΔE_{MM} and solvation free energy ΔG_{SOLV} can be further partitioned into contributions from different interactions, as shown in Eq. (4) and Eq. (5):

$$\Delta E_{MM} = \Delta E_{INT} + \Delta E_{ELE} + \Delta E_{VDW} \quad (4)$$
$$\Delta G_{SOLV} = \Delta G_{GB/PB} + \Delta G_{SA} \quad (5)$$

ΔE_{INT} is the change in internal energies, which includes the bond, angle and dihedral energies. ΔE_{ELE} and ΔE_{VDW} are the electrostatic energies and the van der Waals energies, respectively. ΔE_{ELE} is usually calculated using Coulomb's law with atomic charges from the MM force field, so the values depend on the charges used for the protein and the ligand.

$$\Delta G_{SA} = \gamma SASA + b \quad (6)$$

The solvation free energy ΔG_{SOLV} can be divided into two components: the polar term ($\Delta G_{GB/PB}$ and the nonpolar term ΔG_{SA}. The polar part reflects the electrostatic interaction between the solute and the surrounding dielectric medium, which is evaluated using either the Poisson–Boltzmann (PB) or the Generalized Born (GB) model. Compared with PB, the GB approach is more computationally efficient since it provides an analytical form of ΔG_{SOLV}. The nonpolar contribution arises mainly from the energetic cost of cavity formation in the solvent and from van der Waals contacts between solute and solvent molecules. In practice, this term is usually estimated as being proportional to the solvent-accessible surface area (SASA) of the solute and is expressed by Eq. (6), where γ denotes the surface tension parameter and b is an empirical offset.

Text S4. Implementation details of the model

Our model comprises nine equivariant layers that share feature representations, with each layer implemented as a Transformer containing a 128-dimensional hidden state and 16 attention heads. We adopt a sigmoid β-schedule for atomic coordinates with parameters $\beta_1 = 1\times10^{-7}$ and $\beta_T = 2\times10^{-3}$, and a cosine β-schedule with $s = 0.01$. The diffusion process is

performed over T = 1000 steps. Details of the atomic features for both ligands and protein binding pockets are provided in **Tables S6** and **S7**. During training, the model was optimized with an initial learning rate of 0.01, which decays exponentially with a factor of 0.95 and is bounded below by a minimum learning rate of 1×10^{-5}. Additionally, if the validation loss does not improve for 15 consecutive evaluations, the learning rate is further reduced. Model training was conducted on a single NVIDIA GeForce RTX 4090D GPU with 24 GB of memory.

Algorithm S1 Training Procedure of ExpDiff

Input: Protein-ligand complex $\{P, M, v, b\}_{i=1}^{N}$, neural network ϕ_θ, atom coordinate loss scale γ, atom type loss scale δ, expert align loss scale ε

1. **while** ϕ_θ not coverage **do**
2. Sample diffusion step $t \in \{0, ..., T\}$

   Move the complex to make the center of mass protein atoms zero
3. Compute $\boldsymbol{x_t}$: $x_t = \sqrt{\bar{a}_t}\, x_0 + (1-\bar{x}_t)\varepsilon$, where $\varepsilon \in N(0, I)$
4. Compute $\mathbf{z_t}$:

   $\log c = \log(\bar{a}_t z_0 + (1-\bar{a}_t)/K)$

   $z_t = one\_hot(\arg\max_i[g_i + \log z_i])$, where g ~ Gumble (0, 1)
5. Compute $[\hat{x}_0, \hat{z}_0, \hat{v}, \hat{b}]$: $[\hat{x}_0, \hat{z}_0, \hat{v}, \hat{b}] = \phi_\theta([x_t, z_t], t, P)$
6. Compute the posterior atom types and coordinate
7. Compute weighted loss on atom types, atom coordinate, binding affinities and expert align loss, as shown in main txt Eq. (13)
8. Update θ by minimizing L

Text S4. Success rate

Table S8 shows that ExpDiff achieves a strong balance between molecular validity (0.89) and novelty (0.91). Although its validity is slightly lower than DecompOpt (0.95) and KGDiff (0.91), ExpDiff maintains markedly higher novelty than most methods, indicating better exploration of chemical space while still producing chemically reasonable structures. In contrast, methods like TargetDiff show both lower validity and novelty, suggesting more constrained or less stable generation behavior. Overall, ExpDiff demonstrates competitive chemical correctness with superior ability to generate new structures, reflecting a favorable validity–novelty trade-off among the compared models.

Text S5. Comparison across models with bootstrap confidence intervals

Table S9 summarizes the mean performance and corresponding 95% bootstrap confidence intervals for all models across the evaluated metrics, including Vina Score, Vina Minimize, Vina Dock, QED, and SAS. Several clear patterns emerge from these results.

Across all docking-related metrics (Vina Score, Vina Minimize, and Vina Dock), ExpDiff consistently exhibits substantially lower and more favorable scores than the baseline models. Although the confidence intervals of different models show partial overlap due to the high variance introduced by heterogeneous targets, the interval of ExpDiff is always shifted toward the better-performing region, indicating a consistent trend of improved docking affinity.

For QED, ExpDiff achieves competitive drug-likeness relative to the baselines. Its confidence interval is positioned between those of AR and DeepICL, suggesting that ExpDiff maintains favorable physicochemical properties while prioritizing structural optimization for binding. SAS results show that ExpDiff generates molecules with higher synthetic accessibility scores compared to several baseline methods. While ExpDiff does not match the synthetic ease of models specifically tuned for generating high-SAS molecules, its confidence interval remains within a reasonable range and does not indicate severe penalties in synthetic feasibility.

Overall, the confidence interval analysis demonstrates that ExpDiff provides steady and directional improvements in structure-based metrics while maintaining balanced drug-like and synthetic properties. These patterns support the model's general capability in generating chemically meaningful and structurally optimized molecules across diverse protein targets.

Text S6 Effect size and significance testing across molecular generation metrics

Here we summarize the results presented in Table S10, which reports Cohen's d and paired bootstrap p-values for the comparisons between ExpDiff and the baseline models across the five evaluation metrics.

For the three docking-related metrics (Vina Score, Vina Minimize, and Vina Dock), ExpDiff achieves lower average scores than all baselines, and the corresponding effect sizes are often in the medium to large range. This indicates a clear trend in favor of ExpDiff. At the same time, the bootstrap p-values remain relatively high, and none of the differences are statistically significant after applying the Benjamini–Hochberg procedure. This outcome is not unexpected: docking performance varies considerably from target to target, and such variability directly reduces statistical power, even when the average improvement is noticeable.

For QED, the effect sizes are small and inconsistent in sign across baselines, suggesting that ExpDiff produces molecules with drug-likeness properties similar to those generated by the existing methods. The SAS metric shows a similarly mixed pattern, reflecting the inherent balance between improving binding affinity and maintaining synthetic feasibility.

Taken together, these results show that ExpDiff offers preferable docking performance on average while performing on par with other models in terms of physicochemical and synthetic attributes. The absence of statistical significance reflects the high variability of multi-target evaluations rather than a lack of performance difference. To give a complete picture, we report both effect sizes and hypothesis testing results so that readers can interpret practical trends alongside statistical uncertainty.

Supporting Tables

Table S1. Comparison of molecular geometry and steric clashes between molecules generated by ExpDiff and baseline models. The top-performing results are highlighted in bold, and the second-best results are underlined.

Method	Static Geometry		Clash	
	JSD_{BL}	JSD_{BA}	$Ratio_{cca}$	$Ratio_{cm}$
LIGAN	0.4645	0.5673	**0.0096**	**0.0718**
GrapgBP	0.5182	0.5645	0.8634	0.9974
Pocket2Mol	0.5433	0.4922	0.0576	0.4499
TargetDiff	<u>0.2659</u>	<u>0.3769</u>	0.0483	0.4920
DecopDiff	**0.2576**	**0.3473**	0.0462	0.5248
DeepICL	0.4256	0.4290	0.0938	0.8100
PocketFlow	0.3968	0.3838	0.0494	0.6222
ExpDiff(ours)	0.2808	0.3886	<u>0.0161</u>	<u>0.2696</u>

Table S2. Comparison of bond angles of molecules generated by ExpDiff and baseline models. — denotes absence. - indicates a single bond, = indicates a double bond, and # indicates an aromatic bond in column JSD_{BA}.

JSD_{BA}	Pocket2Mol	TargetDiff	DecompDiff	DeepICL	PocketFlow	ExpDiff(ours)
C#C-C	0.6477	0.6845	0.8174	0.6960	0.7588	0.6667
C-C#N	0.5830	0.7437	0.7254	—	0.7733	0.6935
C-C-C	0.4663	0.2955	0.2306	0.3507	0.3670	0.3035
C-C-N	0.4790	0.2738	0.1987	0.2628	0.2906	0.2955
C-C-O	0.5078	0.3335	0.2124	0.4749	0.3280	0.3778
C-C=C	0.2826	0.1815	0.2215	0.1694	0.1904	0.2032
C-C=N	0.3507	0.2075	0.2094	0.3829	0.2326	0.2662
C-N-C	0.3981	0.2915	0.1952	0.1829	0.2324	0.3074
C-N-N	0.4997	0.2626	0.2825	0.6549	0.2429	0.2861
C-N-O	0.6173	0.3263	0.3120	—	—	0.3998
C-N=C	0.3728	0.3105	0.3467	0.4107	0.3513	0.4334
C-N=N	0.7062	0.4400	0.3917	—	0.5278	0.7117
C-O-C	0.4204	0.2865	0.1882	0.4813	0.2974	0.3067
C-O-N	0.6140	0.4312	0.4064	—	—	0.4426
C=C-N	0.3732	0.2359	0.2574	0.5192	0.2720	0.2424
C=C=C	0.7373	0.7445	0.7703	—	—	0.7376
N#C-C	0.5830	0.7437	0.7254	—	0.7733	0.2662
N-C-N	0.5544	0.3058	0.2994	0.5744	0.2324	0.3349
N-C-O	0.5879	0.3926	0.3029	0.7891	0.3948	0.4253
N-C=N	0.3986	0.2175	0.2593	0.4229	0.2622	0.2321
N-C=O	0.2347	0.2664	0.1197	0.2490	0.3140	0.2900
N-N-O	0.7639	0.5862	0.4831	—	—	0.5806
N=C-N	0.3986	0.2175	0.2593	0.3829	0.5209	0.2321
O=C-N	0.2347	0.2664	0.1197	0.2490	0.3140	0.2900

Table S3. Effect of different alignment loss factor ε on the affinity of ExpDiff generated molecules. Bolding indicates the best affinity value. $ExpDiff_0$ indicates that feature alignment is not applicable.↑ means the higher the better and ↓ means the lower the better.

Method	Vina Score(↓)		Vina Minimize(↓)		Vina Dock(↓)	
	Average	Median	Average	Median	Average	Median
$ExpDiff_0$	-8.041	-8.613	-8.786	-8.842	-9.433	-9.421
$ExpDiff_{0.05}$	-8.278	-8.802	-8.965	-9.029	-9.561	-9.487
$ExpDiff_{0.1}$	-9.096	-9.519	-9.579	-9.708	-9.920	-9.942
$ExpDiff_{0.5}$	-8.724	-9.375	-9.402	-9.064	-9.851	-10.024
$ExpDiff_1$	-8.030	-8.454	-8.681	-8.620	-9.258	-9.204

Table S4. ExpDiff docking results on different protein pockets. ↑ means the higher the better and ↓ means the lower the better. Ave. indicates average. Med. indicates median.

Categorization	PDB ID	Vina Score(↓)		Vina Minimize(↓)		Vina Dock(↓)		SA(↑)	
		Ave.	Med.	Ave.	Med.	Ave.	Med.	Ave.	Med.
Nuclear Receptor	1A52	-12.18	-12.42	-12.47	-12.77	-12.61	-12.86	0.46	0.46
	2PQG	-11.41	-11.57	-11.62	-11.83	-11.79	-12.08	0.54	0.53
Protease	1AQ7	-7.44	-7.45	-8.04	-7.89	-8.42	-8.48	0.67	0.68
	1PPD	-12.88	-12.85	-13.05	-13.01	-13.23	-13.09	0.54	0.55
Epigenetic regulation	4WXX	-11.40	-11.09	-11.68	-11.41	-11.46	-11.50	0.44	0.45
	2HKO	-11.88	-12.16	-12.32	-12.45	-12.82	-12.63	0.53	0.52
Transporter	1B0U	-6.75	-6.96	-6.35	-6.25	-7.18	-7.10	0.44	0.44
	8E56	-11.80	-11.80	-11.95	-12.01	-12.07	-12.04	0.44	0.43

Table S5. The PDB IDs in the test set. All PDBs were subjected to molecular docking experiments, while the PDBs with a gray background were not included in the binding free energy experiments.

Index	PDB ID	Index	PDB ID	Index	PDB ID	Index	PDB ID	Index	PDB ID
0	2z3h	20	1dxo	40	3jyh	60	4pxz	80	3o96
1	4aaw	21	1gg5	41	4iwq	61	2gns	81	4qlk
2	4yhj	22	5q0k	42	1l3l	62	1ai4	82	3hy9
3	14gs	23	5b08	43	5ngz	63	5mma	83	4bel
4	2v3r	24	2azy	44	1e8h	64	2cy0	84	3nfb
5	4rn0	25	5i0b	45	2e24	65	3w83	85	4m7t
6	1fmc	26	1phk	46	2hcj	66	2e6d	86	3u9f
7	3daf	27	4keu	47	3kc1	67	4rv4	87	4aua
8	1a2g	28	4q8b	48	1d7j	68	5d7n	88	2f2c
9	5w2g	29	1djy	49	4ja8	69	5mgl	89	3chc
10	3dzh	30	5l1v	50	4u5s	70	1h36	90	1k9t
11	3g51	31	4zfa	51	4iiy	71	4gvd	91	1h0i
12	1coy	32	2rma	52	3v4t	72	4tos	92	4z2g
13	2jjg	33	3b6h	53	3tym	73	5aeh	93	3af2
14	2rhy	34	2zen	54	4d7o	74	4h3c	94	1jn2
15	2pqw	35	4p6p	55	3ej8	75	4rlu	95	3li4
16	4g3d	36	3u5y	56	1rs9	76	4xli	96	3pnm
17	5bur	37	4f1m	57	4kcq	77	3l3n	97	1afs
18	3gs6	38	4tqr	58	3pdh	78	5tjn	98	4azf
19	1r1h	39	4lfu	59	1umd	79	5liu	99	2pc8

Table S6. Atomic features of ligand used in the model.

Types of atoms	Aromatic	Index
H	False	0
C	False	1
	True	2
N	False	3
	True	4
O	False	5
	True	6
F	False	7
P	False	8
	True	9
S	False	10
	True	11
Cl	False	12

Table S7. Atomic features of pocket used in the model.

Feature	Property	Index
Types of atoms	H, C, N, O, S, Se	0-5
Amino acids	ALA, CYS, ASP, GLU, PHE, GLY, HIS, ILE, LYS, LEU MET, ASN, PRO, GLN, ARG, SER, THR, VAL, TRP, TYR	6-26
Is backbone	Yes or No	27

Table S8. Molecular validity and novelty.

Methods	Valid	Novel
DecompOpt	0.95	0.85
KGDiff	0.91	0.93
ResGen	0.87	0.89
TargetDiff	0.82	0.72
ExpDiff	0.89	0.91

Table S9. Performance comparison across models with bootstrap confidence intervals

Metric	Model	CI_low	CI_high	Metric	Model	CI_low	CI_high
Vina Score	AR	-6.237	-5.202	QED	AR	0.476	0.515
	DeepICL	-4.204	-3.313		DeepICL	0.601	0.619
	ExpDiff	-9.531	-7.387		ExpDiff	0.512	0.550
	KGDiff	-8.765	-6.173		KGDiff	0.528	0.567
	PMDM	1.118	6.127		PMDM	0.545	0.570
	Pocket2mol	-5.708	-4.660		Pocket2mol	0.545	0.576
	PocketFlow	-3.840	-1.877		PocketFlow	0.512	0.528
	TargetDiff	-6.342	-3.698		TargetDiff	0.462	0.504
Vina Minimize	AR	-6.561	-5.798	SAS	AR	6.531	6.786
	DeepICL	-6.192	-5.598		DeepICL	6.997	7.107
	ExpDiff	-9.894	-8.523		ExpDiff	5.047	5.392
	KGDiff	-9.264	-7.962		KGDiff	6.966	7.202
	PMDM	-4.018	-1.197		PMDM	4.866	5.097
	Pocket2mol	-6.962	-5.956		Pocket2mol	6.527	6.739
	PocketFlow	-5.792	-5.064		PocketFlow	2.711	2.859
	TargetDiff	-7.121	-5.750		TargetDiff	6.715	6.916
Vina Dock	AR	-7.128	-6.375				
	DeepICL	-7.484	-6.794				
	ExpDiff	-10.322	-9.116				
	KGDiff	-9.800	-8.655				
	PMDM	-7.528	-6.548				
	Pocket2mol	-7.712	-6.693				
	PocketFlow	-7.383	-6.876				
	TargetDiff	-8.203	-7.367				

Table S10. Effect size and significance testing across molecular generation metrics

Metric	baseline	cohens_d	p_raw	p_fdr
Vina Socre	AR	-0.6459935	0.551	0.551
	DeepICL	-1.1258104	0.5378	0.551
	KGDiff	-0.1585633	0.4556	0.551
	PMDM	-1.2272413	0.4818	0.551
	Pocket2mol	-0.778038	0.5486	0.551
	PocketFlow	-1.0141556	0.4894	0.551
	TargetDiff	-0.5397269	0.4934	0.551
Vina Minimize	AR	-1.0996163	0.5168	0.5322
	DeepICL	-1.2614806	0.516	0.5322
	KGDiff	-0.1821689	0.4622	0.5322
	PMDM	-1.1634747	0.4758	0.5322
	Pocket2mol	-0.920255	0.5322	0.5322
	PocketFlow	-1.3746149	0.4972	0.5322
	TargetDiff	-0.789585	0.4966	0.5322
Vina Dock	AR	-1.1907762	0.5062	0.515
	DeepICL	-1.0565881	0.5044	0.515
	KGDiff	-0.1636335	0.5046	0.515
	PMDM	-0.9728232	0.49	0.515
	Pocket2mol	-0.9026164	0.5038	0.515
	PocketFlow	-1.1294487	0.5096	0.515
	TargetDiff	-0.7430722	0.515	0.515
QED	AR	0.36715085	0.4842	0.5104
	DeepICL	-1.0336063	0.5104	0.5104
	KGDiff	-0.1678653	0.4996	0.5104
	PMDM	-0.3294221	0.509	0.5104
	Pocket2mol	-0.3290686	0.5042	0.5104
	PocketFlow	0.14646568	0.5058	0.5104
	TargetDiff	0.46866532	0.4844	0.5104
SAS	AR	-1.861642	0.501	0.5108
	DeepICL	-2.8317657	0.4928	0.5108
	KGDiff	-2.5015834	0.4982	0.5108
	PMDM	0.32942212	0.509	0.5108
	Pocket2mol	-1.9531672	0.4968	0.5108
	PocketFlow	3.60951017	0.5108	0.5108
	TargetDiff	-2.2383085	0.4974	0.5108

Supporting Figures

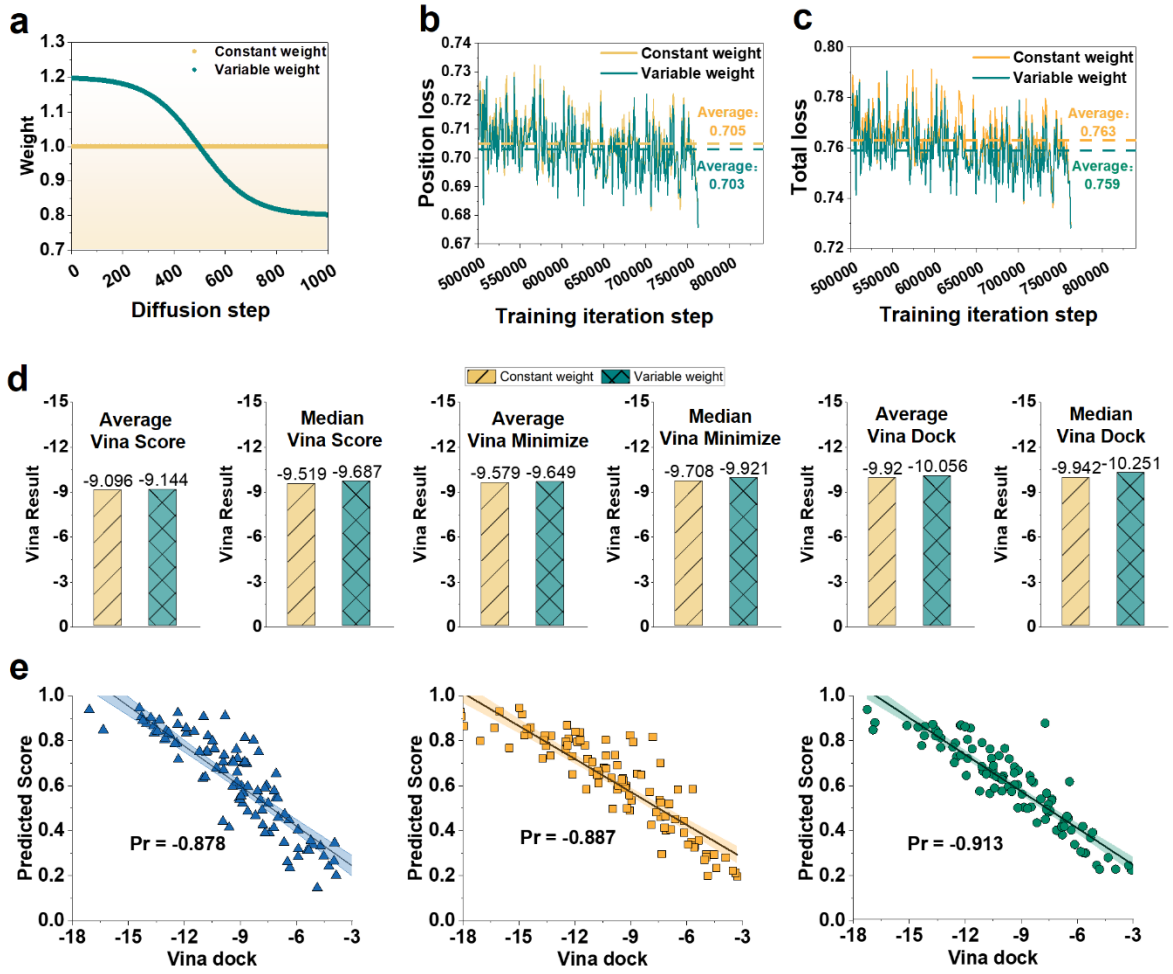

Figure S1. Effect of using constant and variable weights on training ExpDiff. **a** denotes the weight values. **b** and **c** denote the positional and total losses of small molecules, respectively. **d** denotes the docking results. **e** denotes the Pearson correlation (Pr) between the affinity prediction network and the Vina dock results. Blue color indicates models without alignment, green color indicates models with an alignment factor of 0.1, and yellow color indicates that variable weights were applied.

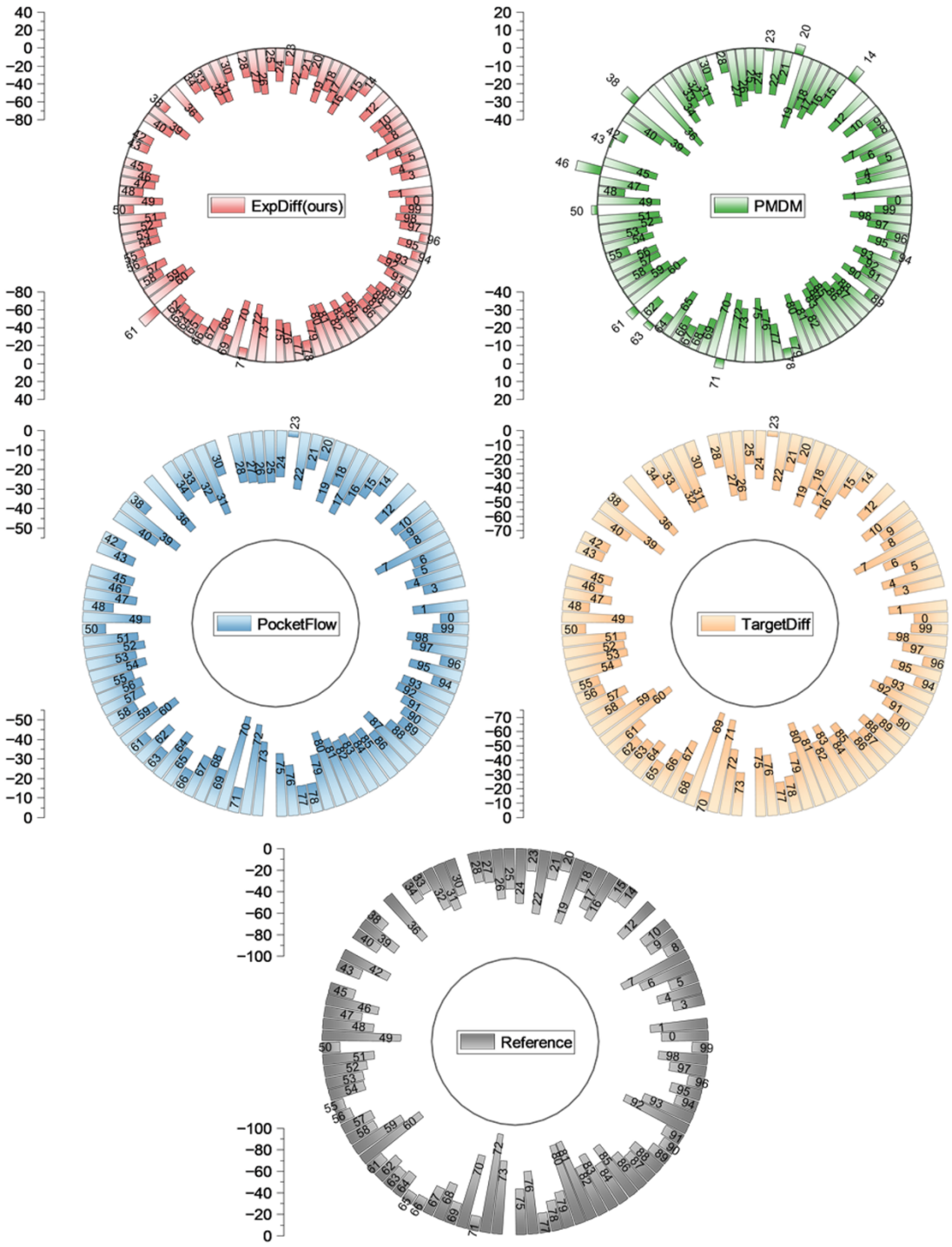

Figure S2. The binding free energy of ExpDiff and the baseline models.

Reference

[1] Lin H, Zhao G, Zhang O, et al. CBGBench: Fill in the Blank of Protein-Molecule Complex Binding Graph[C]. The Thirteenth International Conference on Learning Representations (ICLR), **2025**.

[2] Ramachandran S, Kota P, Ding F, et al. Automated minimization of steric clashes in protein structures[J]. Proteins: Structure, Function, and Bioinformatics, **2011**, 79(1): 261-270.

[3] Yang M, Bo Z, Xu T, et al. Uni-GBSA: an open-source and web-based automatic workflow to perform MM/GB(PB)SA calculations for virtual screening[J]. Briefings in Bioinformatics, **2023**; 24(4): bbad218.

[4] Eastman P, Friedrichs M S, Chodera J D, et al. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation[J]. Journal of Chemical Theory and Computation, **2013**; 9(1): 461-469.

[5] Honig B, Gilson M. Calculation of the total electrostatic energy of a macromoleular system: solvation energies, binding energies, and conformational analysis. Proteins **1988**; 4:7–18

[6] Wang J, Hou T, Xiaojie X. Recent advances in free energy calculations with a combination of molecular mechanics and continuum models. Curr Comput Aided Drug Des **2006**; 2(3): 287–306.

[7] Wang E, Hou T, Zhang JZH, et al. . End-point binding free energy calculation with mm/pbsa and mm/gbsa: strategies and applications in drug design. Chem Rev **2019**; 119(16): 9478–508.